

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:28 ; Search time 170.72 Seconds
(without alignments)
17.102 Million cell updates/sec

Title: US-09-331-631a-3_COPY_74_116

Perfect score: 250

Sequence: 1 NQDPQPTDCQQCORRCRQE..... RQQYCQRRCKEICEEEBYY 43

Scoring table: BLOSUM62

GapOp 10.0 , GapExt 0.5

Searched: 195891 seqs, 6790655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR:66:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Match | Length | DB ID | Description |
|------------|-------|-------|-------|--------|--------|-------------------------|
| 1 | 112 | 44.8 | 566 | 2 | S22477 | vicilin precursor |
| 2 | 102 | 40.8 | 509 | 2 | S00059 | alpha-globulin type B p |
| 3 | 100 | 40.0 | 588 | 1 | F9WNAB | hypothetical protein |
| 4 | 92 | 36.8 | 411 | 2 | T23475 | alpha-globulin type B p |
| 5 | 90 | 36.0 | 605 | 2 | S06398 | hypothetical protein |
| 6 | 78 | 31.2 | 572 | 2 | T29880 | hypothetical protein |
| 7 | 77.5 | 31.0 | 425 | 2 | T28592 | hypothetical protein |
| 8 | 77.5 | 31.0 | 600 | 2 | T28593 | hypothetical protein |
| 9 | 77 | 30.8 | 810 | 2 | T24430 | protein PV100 [Imp |
| 10 | 75 | 30.0 | 47 | 2 | JCC557 | arginine/glutamate |
| 11 | 75 | 30.0 | 154 | 2 | T24967 | hypothetical protein |
| 12 | 73 | 29.4 | 242 | 2 | T29699 | hypothetical protein |
| 13 | 73 | 29.2 | 111 | 2 | T20452 | hypothetical protein |
| 14 | 72 | 28.8 | 330 | 2 | T23169 | hypothetical protein |
| 15 | 71 | 28.4 | 388 | 2 | T21887 | hypothetical protein |
| 16 | 71 | 28.4 | 388 | 2 | T21888 | hypothetical protein |
| 17 | 71 | 28.4 | 438 | 2 | T21889 | hypothetical protein |
| 18 | 71 | 28.4 | 445 | 2 | T21889 | hypothetical protein |
| 19 | 69.5 | 27.8 | 335 | 2 | T21560 | hypothetical protein |
| 20 | 69.5 | 27.8 | 335 | 2 | T21561 | hypothetical protein |
| 21 | 67.5 | 27.0 | 314 | 2 | T27686 | hypothetical protein |
| 22 | 67.5 | 27.0 | 335 | 2 | T21559 | hypothetical protein |
| 23 | 67 | 26.8 | 1513 | 2 | T23681 | hypothetical protein |
| 24 | 66.5 | 26.6 | 342 | 2 | JCC110 | brain-specific mean |
| 25 | 66.5 | 26.6 | 637 | 2 | S05221 | globulin Bg1 pre |
| 26 | 26.4 | 26.4 | 165 | 2 | T20468 | hypothetical protein |
| 27 | 64.5 | 25.8 | 654 | 2 | T20136 | hypothetical protein |
| 28 | 63.5 | 25.4 | 33 | 2 | A1822 | anticancer pept |
| 29 | 63.5 | 25.4 | 273 | 2 | T20930 | hypothetical protein |

ALIGNMENTS

| RESULT | 1 | Score | Match | Length | DB ID | Description |
|---|---|--|-------|--------|-------|---|
| S22477 | | | | | | vicilin precursor - cacao |
| C-Species: Theobroma cacao (cacao) | | | | | | C-Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999 |
| C-Accession: S22477; S22478; S18105; S22050 | | | | | | R-McHenry, L.; Fritz, P.J. |
| Plant Mol. Biol. 18, 1173-1176, 1992 | | | | | | A>Title: Comparison of the structure and nucleotide sequence of vicilin genes of caco |
| A:Reference number: S22477; MUID:92288309 | | | | | | A:Accession: S22477 |
| A:Molecule type: DNA | | | | | | A:Residues: 1-566 <MOH> |
| A:Cross-references: EMBL:X62625 | | | | | | A:Accession: S22478 |
| A:Molecule type: mRNA | | | | | | A:Residues: 1-452 <MO2> |
| A:Cross-references: EMBL:X62626 | | | | | | A:Cross-references: EMBL:X62625 |
| C:Genetics: | | | | | | C:Keywords: seed; storage protein |
| A:Introns: 211/1; 269/3; 296/3; 391/3; 502/1 | | | | | | F:1-24 /Domain: signal sequence #status predicted <SIG> |
| C:Superfamily: glycinin | | | | | | F:25-566 /Product: vicilin #status predicted <MAT> |
| C:Keywords: seed; storage protein | | | | | | Query Match |
| F:1-24 /Domain: signal sequence #status predicted <SIG> | | | | | | Best Local Similarity 44.8%; Score 112; DB 2; Length 566; |
| F:25-566 /Product: vicilin #status predicted <MAT> | | | | | | Matches 19; Conservative 10; Mismatches 11; Indels 0; Gaps 0; |
| Query | 2 | ODDPQPDCQQCORRCRQEESPRQDQYCQRRCKEICEEE | 41 | Db | 78 | PEELORQYQOCQGRQEQQQOREQQCQRCWEQKQE 117 |

RESULT

2

S08059

alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)

N:Alternate names: seed storage protein

C:Species: *Gossypium hirsutum* (upland cotton)

C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993

C:Accession: S08059

R-Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.

Plant Mol. Biol. 9, 533-546, 1987

A>Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.

A:Reference number: S06398

A:Accession: S08059

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-509 <CIL>

C:Superfamily: glycinin

Query Match 40.8%; Score 102; DB 2; Length 509;
 Best Local Similarity 48.7%; Pred. No. 0.001; Indels 2; Gaps 1;
 Matches 19; Conservative 8; Mismatches 10; Indels 2; Gaps 1;
 Matches 19; Conservative 8; Mismatches 10; Indels 2; Gaps 1;

Qy 4 DPOTOCQCQCORRCRQESGRQQYCORRKEICEEEE 42
 Db 1 DPQRYYECQECRQEE - RQQPOCQQRCLKRFEEQQ 37

RESULT 3
 FWCNAB
 alpha-globulin B precursor (clone C72) - upland cotton
 N;Alternate names: seed storage protein; vicilin precursor
 C;Species: Gossypium hirsutum (upland cotton)
 C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
 C;Accession: A30838; S05911
 R;Chian, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.
 Plant Mol. Biol. 7, 475-489, 1986
 A;Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII. C
 A;Reference number: A30838
 A;Accession: A30838
 A;Molecule type: mRNA
 A;Residues: 1-588 <CHL>
 A;Cross-references: GB:MB6911; NTDB:9167374; PIDN:AAA33071.1; PID:9167375
 A;Experimental source: var. Coker 201
 R;Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
 Plant Mol. Biol. 9, 533-546, 1987
 A;Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. C
 A;Reference number: S06398
 A;Accession: S06398
 A;Status: not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-605 <CHL>
 C;Superfamily: glycinin
 F;1-24;Domain: signal sequence #status predicted <SIG>
 F;25-605;Product: alpha-globulin type A #status predicted <MAT>
 F;417;Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 40.0%; Score 100; DB 1; Length 588;
 Best Local Similarity 45.0%; Pred. No. 0.0019; Indels 2; Gaps 1;
 Matches 18; Conservative 10; Mismatches 10; Indels 2; Gaps 1;

Qy 3 DDPQPTDCQCQCORRCRQESGRQQYCORRKEICEEEE 42
 Db 81 EDPQRYYECQECRQEE - RQQPOCQQRCLKRFEEQQ 118

RESULT 4
 T29475
 hypothetical protein T01D1.6 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C;Accession: T29475
 R;Bradshaw, H.; Wohldmann, P.
 submitted to the EMBL Data Library, November 1996
 A;Description: The sequence of *C. elegans* cosmid T01D1.
 A;Reference number: Z20623
 A;Accession: T29475
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-411 <BRA>
 A;Cross-references: EMBL:U80455; PIDN: AACB37887.1; GSPDB:GN00020; CESP:T01D1.6
 A;Experimental source: strain Bristol NZ; clone T01D1
 A;Gene: CESP:T01D1.6
 A;Map position: 2
 A;Introns: 25/3; 304/3
 C;Superfamily: gladin

Query Match 36.0%; Score 92; DB 2; Length 411;
 Best Local Similarity 39.5%; Pred. No. 0.0059; Indels 4; Gaps 2;
 Matches 17; Conservative 9; Mismatches 13; Indels 4; Gaps 2;

Qy 2 0DPDQPTDCQCQCORRCRQESGRQQYCORRKEICEEEE 43
 Db 138 QQPAAQOCQOCQACPQQO--QPPQQCQOCQCTQSDQY 177

RESULT 5
 S06398
 alpha-globulin type A precursor - upland cotton
 N;Alternate names: seed storage protein
 C;Species: Gossypium hirsutum (upland cotton)
 C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
 C;Accession: S06398
 R;Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
 Plant Mol. Biol. 9, 533-546, 1987
 A;Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. C
 A;Reference number: S06398
 A;Accession: S06398
 A;Status: not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-605 <CHL>
 C;Superfamily: glycinin
 F;1-24;Domain: signal sequence #status predicted <SIG>
 F;25-605;Product: alpha-globulin type A #status predicted <MAT>
 F;417;Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 36.0%; Score 90; DB 2; Length 605;
 Best Local Similarity 35.9%; Pred. No. 0.021; Indels 0; Gaps 0;
 Matches 14; Conservative 10; Mismatches 15; Indels 0; Gaps 0;

Qy 2 0DPDQPTDCQCQCORRCRQESGRQQYCORRKEICEEEE 40
 Db 115 QQPDPQFQKRCQRCOWEQOPERKQOCVRCREQYQED 153

RESULT 6
 T29880
 hypothetical protein F19G12.7 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C;Accession: T29880
 R;Nhan, M.; Le, T.T.
 submitted to the EMBL Data Library, April 1996
 A;Description: The sequence of *C. elegans* cosmid F19G12.
 A;Reference number: Z20704
 A;Accession: T29880
 A;Status: Preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-572 <NTA>
 A;Cross-references: EMBL:U51997; PIDN: AAC48159.1; GSPDB:GN00028; CESP:F19G12.7
 A;Experimental source: strain Bristol NZ; clone F19G12
 A;Genetics:
 A;Gene: CESP:F19G12.7
 A;Map position: X
 A;Introns: 18/3

Query Match 31.2%; Score 78; DB 2; Length 572;
 Best Local Similarity 35.8%; Pred. No. 0.36; Indels 20; Gaps 4;
 Matches 19; Conservative 7; Mismatches 7; Indels 20; Gaps 4;

Qy 5 PQTDCQCQCORRCRQ-----QESGRQQYCORRKE--ICEEEE 40
 Db 380 PQ - CQQCQNCQQFAPVCBQQCASICLHQPSAPCQQ - GNTCQQFAPVCSEQ 429

RESULT 7
 T18592
 hypothetical protein AC3.3 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T18592
R;McMurray, A.
A;Reference number: 218995
A;Accession: T18592
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-425 <WIL>
A;Cross-references: EMBL:Z71177; PIDN:CAA94867.1; GSPDB:GN00023; CESP:AC3.3
A;Experimental source: clone AC3
C;Genetics:
A;Gene: CESP:AC3.3
A;Map position: 5
A;Introns: 18/3
C;Superfamily: giladin

Query Match 31.0%; Score 77.5; DB 2; Length 425;
Best Local Similarity 40.0%; Pred. No. 0; 33; Mismatches 15; Indels 0; Gaps 0;
Matches 18; Conservative 9; Mismatches 7; Indels 11; Gaps 4;
Qy 6 QPTDCQ---QCORRCCR---QOESGPROQQYCQRRCKE---ICEEE 40
Db 236 OTTCQOAVPOCQOCAPOCQQSAPOCQQ-CONTCOQAAAPVCOQQ 279

RESULT 8

T18593 hypothetical protein AC3.4 - *Caenorhabditis elegans*
Species: *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Accession: T18593
R;McMurray, A.
submitted to the EMBL Data Library, April 1996

A;Accession: T18593
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-600 <WIL>
A;Cross-references: EMBL:Z71177; PIDN:CAA94868.1; GSPDB:GN00023; CESP:AC3.4
C;Genetics:
A;Experimental source: clone AC3
A;Gene: CESP:AC3.4
A;Map position: 5
A;Introns: 292/2; 374/3; 572/2

Query Match 31.0%; Score 77.5; DB 2; Length 600;
Best Local Similarity 40.0%; Pred. No. 0; 42; Mismatches 9; Indels 11; Gaps 4;
Matches 18; Conservative 9; Mismatches 7; Indels 11; Gaps 4;
Qy 6 QPTDCQ---QOQRCCR---QOESGPROQQYCQRRCKE---ICEEE 40
Db 103 OTTCQOAVPOCQOCAPOCQQSAPOCQQ-CQNTCQAAAPVCOQQ 146

RESULT 9

T4430 protein PV100 [imported] - winter squash
Species: *Cucurbita maxima* (winter squash)
C;Species: *Cucurbita maxima* (winter squash)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T4430
R;Yamada, K.; Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.
J. Biol. Chem. 274, 2563-2570, 1999
A;Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a single
A;Reference number: 222767; MUID:99107919
A;Accession: T4430
A;Status: preliminary; translated from GB/EMBL/DDBJ

Query Match 30.0%; Score 75; DB 2; Length 154;
Best Local Similarity 41.9%; Pred. No. 0; 28; Mismatches 13; Indels 0; Gaps 0;
Matches 13; Conservative 3; Mismatches 15; Indels 0; Gaps 0;
Qy 3 DDPQTCQCCQRCRQESGPROQQYCQRRC 33
Db 69 DDPNTPCTQYTSILCSKAYTPPLQQCPKTC 99

RESULT 10

T29699 hypothetical protein F31A3.1 - *Caenorhabditis elegans*
Species: *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
A;Cross-references: EMBL:AB019195; PIDN:BA34056.1; PID:93808062

Query Match 30.8%; Score 77; DB 2; Length 810;
Best Local Similarity 34.2%; Pred. No. 0; 59; Mismatches 13; Indels 0; Gaps 0;
Matches 13; Conservative 10; Mismatches 15; Indels 0; Gaps 0;
C;Accession: JC5557
R;Kimura, M.; Park, S. S.; Sakai, R.; Yamasaki, N.; Funatsu, G.
Biosci. Biotechnol. Biochem. 61, 984-988, 1997
A;Title: Primary structure of arginine/glutamate-rich polypeptide from the seeds
A;Reference number: JC5557; MUID:97357433
A;Accession: JC5557
A;Molecule type: protein
A;Residues: 1-47 <WIL>
A;Experimental source: seed
C;Comment: This protein is a storage protein which provides nitrogen and carbon reser
F,12-33,16-29/Disulfide bonds: #status predicted

C;Accession: T29699
 R;Murray, J.; Le, T.
 Submitted to the EMBL Data Library, May 1996
 A;Description: The sequence of *C. elegans* cosmid F31A3.
 A;Reference number: Z20667
 A;Accession: T29699
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-242 <MUR>
 A;Cross-references: EMBL:U58742; PIDN:AB36856.1; GSPDB:GN00028; CESP:F31A3.1
 A;Experimental source: strain Bristol N2; clone F31A3
 C;Genetics:
 A;Gene: CESP:F31A3.1
 A;Map position: X
 A;Introns: 18/3; 160/3
 C;Superfamily: ultra-high-sulfur keratin

RESULT 13
 T20452
 hypothetical protein E04D5.4 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C;Accession: T20452
 R;McMurray, A.
 submitted to the EMBL Data Library, October 1995
 A;Reference number: Z19377
 A;Accession: T20452
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-111 <AF1>
 A;Cross-references: EMBL:AF016451; PIDN:AB66001.1; GSPDB:GN00023; CESP:C03A7.4
 A;Experimental source: strain Bristol N2; clone C03A7
 A;Gene: CESP:C03A7.4
 A;Map position: 5
 A;Introns: 75/3
 C;Superfamily: gliadin

Query Match 29.4%; Score 73.5; DB 2; Length 242;
 Best Local Similarity 33.3%; Pred. No. 0.57;
 Matches 17; Conservative 10; Mismatches 13; Indels 11; Gaps 4;

Qy 2 QDPDPDCCOCRCR-----QDESGPR-QQY---CORRKEICEEE 42
 Db 77 QQPAAPQOCQCNQTCQOSAPVCQOCAPQOCQFAPACQ-QCNSCQHQ 266

RESULT 14
 T25169
 hypothetical protein T23FL6 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C;Accession: T25169
 R;Wilkinson, J.
 submitted to the EMBL Data Library, October 1996
 A;Reference number: Z19990
 A;Accession: T25169
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-330 <WIL>
 A;Cross-references: EMBL:Z81129; PIDN:CAB03405.1; GSPDB:GN00023; CESP:T23FL6
 A;Experimental source: clone T23FL1
 C;Genetics:

A;Gene: CESP:T23FL6
 A;Map position: 5
 A;Introns: 16/3
 C;Superfamily: gliadin

Query Match 28.8%; Score 72; DB 2; Length 330;
 Best Local Similarity 29.5%; Pred. No. 1;
 Matches 13; Conservative 10; Mismatches 9; Indels 12; Gaps 2;

Qy 7 TDCCOCQRCRQQ-----ESPRQQO--YCQRROKEICE 38
 Db 230 TNCQQCQNSCQNSNTQQTITYWQASPTQSCVPOCQQCQ 273

RESULT 15
 T31887
 hypothetical protein C03A7.4 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
 C;Accession: T31887
 R;Greco, T.; Bradshaw, H.; Elliott, G.
 submitted to the EMBL Data Library, July 1997
 A;Description: The sequence of *C. elegans* cosmid C03A7.
 A;Reference number: Z21096
 A;Accession: T31887
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-388 <GFE>
 A;Cross-references: EMBL:AF016451; PIDN:AB66001.1; GSPDB:GN00023; CESP:C03A7.4
 A;Experimental source: strain Bristol N2; clone C03A7
 A;Gene: CESP:C03A7.4
 A;Map position: 5
 A;Introns: 75/3
 C;Superfamily: gliadin

Query Match 28.4%; Score 71; DB 2; Length 388;
 Best Local Similarity 32.5%; Pred. No. 1.5;
 Matches 13; Conservative 10; Mismatches 11; Indels 6; Gaps 2;

Qy 6 OTDCOC-OCORRC-----RQESGPRQQYCORRKEICE 39
 Db 194 QAOCPQOCQOCQSSCQSSNQCEPACNTQCSDICQ 233

Search completed: March 1, 2001, 15:52:29
 Job time: 553 sec